THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Atomic scale modeling of ordering phenomena in inorganic clathrates

Mattias Ångqvist

Department of Physics
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2018

Atomic scale modeling of
ordering phenomena in inorganic clathrates
Mattias Ångqvist

Department of Physics
Chalmers University of Technology
se-412 96 Göteborg, Sweden
Telephone +46 31 772 10 00

Cover: 3D-rendering of a type I clathrate. The rendering was created in Blender.

Chalmers reproservice
Göteborg, Sweden 2018

Atomic scale modeling of ordering phenomena in inorganic clathrates

Mattias Ångqvist
Department of Physics
Chalmers University of Technology

Abstract

Ordering phenomena in materials often have a crucial impact on materials properties. They are governed by the competition between entropy and energy. Accordingly simulating these aspects requires the construction of models that enable an computationally efficient exploration of the relevant configuration space. Alloy cluster expansions are a technique that is particular suitable for this task as they can be trained to reach high accuracy while being computationally suitable for rapid sampling via Monte Carlo techniques.

In this thesis alloy cluster expansions have been applied in combination with Monte Carlo simulations to study the ordering behavior in various inorganic clathrates. Inorganic clathrates constitute a class of systems with a cage-like framework that can trap loosely bound atoms or even small molecules. These systems are small band gap semiconductors and have a very low lattice thermal conductivity, which gives rise to very good thermoelectric properties. Additionally the host atoms and cage framework can be occupied by a wide range of elements which provides extensive opportunities for property optimization. Inorganic clathrates are thus good examples for systems with a high degree of variability in composition, for which ordering phenomena play a crucial role.

In paper I we studied the ordering behaviour of $Ba_8Ga_{16}Ge_{30}$. Configurations representative for different annealing temperatures were extracted from Monte Carlo simulations and further analyzed to obtain the temperature dependency of the thermoelectric power factor. These data was subsequently used to construct a cluster expansion for the power factor itself, which enabled us to optimize the chemical ordering that maximizes this property. The approach developed in this work is generalizable and can be adapted to other materials.

In paper II we studied the ordering behavior and related properties in the clathrate systems $Ba_8Al_xSi_{46-x}$, $Ba_8Al_xGe_{46-x}$, $Ba_8Ga_xGe_{46-x}$, and $Ba_8Ga_xSi_{46-x}$ as a function of composition and temperature. We achieved very good agreement with the available experimental data for the site occupancy factors (SOFs). This enabled us to reconcile experimental data from different sources and explain the non-monotonic variations of the SOFs. In particular, we provided a rationale for the extreme SOF behavior with varying composition observed in Al based clathrates.

Keywords: Cluster expansion, Monte Carlo, Inorganic clathrates, ordering phenomena

LIST OF PUBLICATIONS

This thesis consists of an introductory text and the following papers:

I **Optimization of the Thermoelectric Power Factor: Coupling between Chemical Order and Transport Properties**
Mattias Ångqvist, Daniel O. Lindroth and Paul Erhart
Chemistry of Materials **28**, 6877 (2016)

II **Understanding Chemical Ordering in Intermetallic Clathrates from Atomic Scale Simulations**
Mattias Ångqvist and Paul Erhart
Chemistry of Materials **29**, 7554 (2017)

Specification of the authors contribution to the publications:

I The author developed the software for constructing and sampling cluster expansions used in the paper. The author constructed the cluster expansions, carried out the Monte Carlo simulations and the associated analysis, and created all figures related to Monte Carlo and cluster expansions.

II The author developed the software for constructing and sampling cluster expansions used in the paper. The author constructed the cluster expansions, carried out the Monte Carlo simulations and associated analysis, and prepared all figures related to Monte Carlo and cluster expansions. The author also wrote the first draft of the paper.

Contents

Glossary

bcc body-centered cubic. 12, 13

BTT Boltzmann transport theory. 34

CE cluster expansion. 5, 11, 14, 15, 17, 19, 30, 32–34

CS compressive sensing. 11, 17, 18

CV cross-validation. 19

DFT density functional theory. 5, 33, 34

LOO-CV leave-one-out CV. 20

MC Monte Carlo. 23, 25, 27, 29, 30, 32–34

SOF site occupancy factor. 9, 10, 33, 34

1 Introduction

1.1 Ordering phenomena

This thesis deals with ordering phenomena in inorganic clathrates. The ordering in a material is dictated by the energetics and the entropy of the material. Common methods used to model mixtures in multi-component systems is demonstrated in this chapter. It will be shown that in order to model the ordering of complex materials the best available models are required.

1.1.1 Entropy and order

Entropy is a measure of the disorder in a system. The second law of thermodynamics states that disorder in a closed system can only increase; in other words closed systems tend to evolve from ordered to disordered states. A simplified example of this is the process of shuffling a fresh deck of cards. The initial state when the deck of cards are in perfect order will "never" appear again in the process of the shuffling since, assuming random shuffling, the probability to end up in the initial state are roughly 1 in 10^{68}. As the shuffling continues the deck of cards becomes more and more disordered.

Another example is the tossing of coins. Imagine having 100 coins and tossing them all at once. A particular sequence can be denoted as head–tail–head–head–… and so on. The probability that all come up heads is the probability of the first coin coming up as heads times the probability that the second coin comes up as heads and so on, hence the probability is $(1/2)^{100}$. This probability is the same for any sequence of the coin toss. Yet from intuition we would expect that all coins coming up as heads should almost never happen.

The resolution to this conundrum lies in the distinction of a specific sequence of coin tosses, called microstates, and the total number of heads and tails of a particular sequence, called a macrostate. All microstates are equally probable but the probabilities for different macrostates vary over a wide range. The probability of the macrostate is the number of all microstates that correspond to that macrostate divided by the number of all possible microstates. The number of total microstates are 2^{100} and the number of microstates for a macrostate is given by $\binom{100}{n} = 100!\,/n!\,(100-n)!$. Consequently the probability of all coins ending up as heads is $1/2^{100}$ whereas the probability of ending up with 50 heads is 10^{29} more likely with a probability of about $10^{29}/2^{100}$. Note that the number of microstates for a particular macrostate is commonly referred to as the multiplicity of that macrostate.

Mathematically entropy is measured as $S = k_B \ln \Omega$, where S is the entropy, k_B is the Boltzmann constant and Ω is the multiplicity of the system. Hence increasing the entropy, or the disorder, simply implies that the system has changed to a more likely state.

The principles of the coin toss example are easily extended to atomic systems. Imagine a system comprised of 100 A atoms. By mixing one B atom into the system the number of possible microstates increases by a factor of 100, as there are 100 possible sites to insert the B atom. A system of atoms obeys the same combinatorics as the coin flipping. Hence the entropy increases by mixing and if we are neglecting the atomic interactions the system will have a tendency to spontaneously start mixing.

1.1.2 Free energy

Entropy goes a long way in explaining the ordering behaviour of materials. Yet, in nature many materials are observed to exhibit ordered states, which have (much) lower entropy than disordered states. Thus if entropy is one half of the picture to explain disordering, the energy is the other half that has to be included. Consider again the case of the AB atomic system, for which we saw that the entropy can be tremendously increased by mixing A and B atoms. In general there is an energetic cost associated with mixing. If the formation of $A-B$ bonds is energetically unfavorable compared to $A-A$ and $B-B$ bonds there is a penalty for mixing and less mixing is expected. On the other hand if $A-B$ bonds are more favorable, the system can both lower its energy and increase its entropy by mixing and more mixing is expected. This interplay of entropy and energy is expressed in the Helmholtz free energy of the system

$$F = U - TS, \tag{1.1}$$

where U is the internal energy and T is the temperature. Generally speaking a system described in the canonical ensemble will strive to minimize its free energy. According to Eq. (1.1) the entropy term becomes more important for higher temperatures. Hence, at a low temperature a system is more likely to be observed in its low energetic state, where the chemical bonds dictate the ordering of the material. As the temperature is increased, however, the material becomes more disordered.

1.1.3 Examples

Many materials can exhibit some form of chemical order which are crucial to understanding their properties. One example is Zn_4Sb_3 which is a material with a high thermoelectric figure of merit, which is partially due to its low thermal conductivity [1]. It has been identified that the source of this low thermal conductivity is due to Zn disorder and Zn nanostructuring which are sources of phonon scattering [2, 3, 4].

Skutterudites are another example of material with chemical order, which are a type of cage like materials which are high performance thermoelectric materials [5, 6, 7, 8, 9]. The skutterudites owe some of its success as good thermoelectric materials from their low thermal conductivity. These materials can lower their thermal conductivity by filling these voids with rare earth metals [10, 11]. It has been observed that the minima of the thermal conductivity is obtained for a partial filling of the voids for some skutterudites [7]. An understanding of how these filler atoms order can thus be very helpful in order to design optimal skutterudites [12].

Yet another example of thermoelectric materials with chemical ordering are the inorganic clathrates. The clathrates possesses a large unit cell which can support a wide range of elemental components and compositions. The ordering in inorganic clathrates show large variations with compositions and constituent atoms [13, 14, 15]. The power factor, which is an integral part of thermoelectric efficiency, has been demonstrated to be able to be optimized by more than 60% for certain chemical ordering patterns [16]. Inorganic clathrates are described in more detail in chapter 2.

1.2 Mixing in multi-component systems

1.2.1 The dilute limit

Consider a system of N particles where all particles are of type A except for n particles of type B. This system is said to be in the dilute limit if $N \gg n$. In the dilute limit the B particles are so few and spread out so it is assumed that interactions between B particles can be neglected. Hence, ΔF, the free energy associated with adding or subtracting a B particle, is constant as a function of

composition. The equilibrium concentration in the dilute limit is then given by $c = c_0 \exp\left[-\Delta F / k_B T\right]$, where c_0 is the concentration of sites available for substitution.

1.2.2 Concentrated solutions: Mean-field treatment

When the condition $N \gg n$ is no longer fulfilled the system is said to be in the concentrated limit. Here, the B atoms have become so numerous that their mutual interaction must be taken into account and the free energy of adding or removing a B atom becomes dependent of the concentration. The simplest treatment applied in the concentrated limit is provided by the mean field approximation to the Ising model. The energy, entropy and the free energy of the system can be expressed in terms of an order parameter, which describes, e.g., the average number of $A-B$ bonds in the structure. The order parameter observed for a certain temperature and concentration is the one that minimizes F. The Helmholtz free energy of the system is given by [17, 18]

$$\begin{aligned} \Delta F_{\text{mix}} &= \Delta U_{\text{mix}} - T\Delta S_{\text{mix}}, \\ \Delta U_{\text{mix}} &= \omega c(1-c), \\ \Delta S_{\text{mix}} &= -k_B \left[c \ln c + (1-c)\ln(1-c)\right], \end{aligned} \tag{1.2}$$

where ω describes the energy cost of mixing and creating an $A-B$ bond, c is the B concentration and k_B is the Boltzmann constant. The effects of ΔU_{mix} and T on ΔF_{mix} are illustrated in Fig. 1.1. Some comments are in order. First, all expressions are symmetric in the concentration c. Secondly, the mixing entropy, ΔS, is positive across the entire concentration range. Consequently the contribution of entropy to the free energy is always negative. Hence, for $T \to \infty$, entropy will be the dominating term and the free energy will always be minimized by mixing ($\Delta F_{\text{mix}} < 0$. For the case of $\Delta U_{\text{mix}} > 0$ there is a competition between energy and entropy that determines whether the components mix or segregate. For $\Delta U_{\text{mix}} > 0$ and low T, ΔF_{mix} can change sign, giving rise to a miscibility gap (Fig. 1.1d).

This very simple form is not suitable for modeling the vast majority of systems. It can, however, be generalized leading to the semi-empirical CALPHAD approach to modeling phase diagrams. In this case, the various contributions to the mixing energy and entropy are expressed in polynomial expansions in temperature and composition. The expansion coefficients are most commonly obtained by fitting to experimental data and, more recently, also from first-principles calculations. The resulting models are widely used in industry for alloy design and optimization.

1.2.3 Concentrated solutions: Beyond mean-field

The mean-field approach outlined in the previous section is fundamentally semi-empirical in its nature and requires input in the form of either experimental or

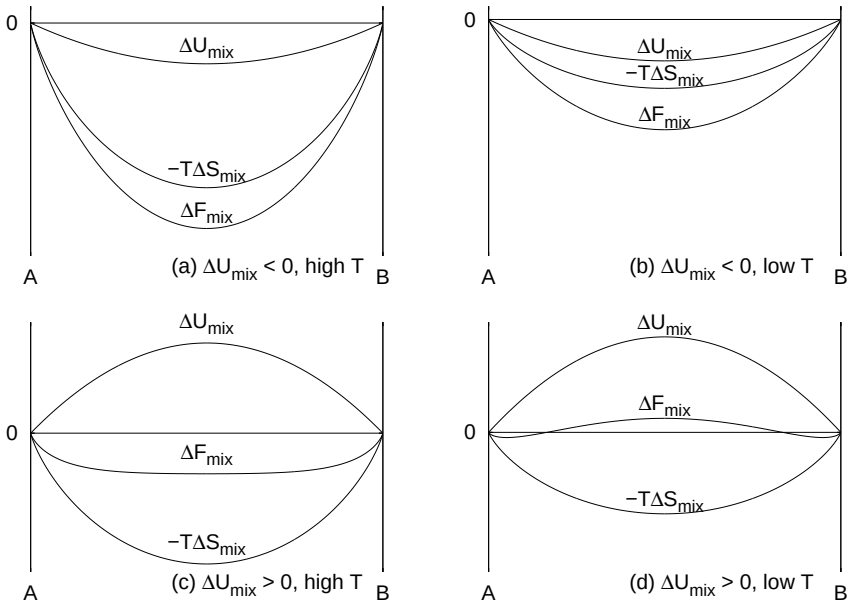

Figure 1.1: The free energy of mixing ΔF_{mix} as a function of composition assuming mixing is (a,b) energetically favorable ($\Delta U_{\mathrm{mix}} < 0$) and (c,d) unfavorable ($\Delta U_{\mathrm{mix}} > 0$), respectively, at (a,c) high and (b,d) low temperature T.

computational data. Experimental data is often difficult and/or expensive to acquire; there are also various cases, in which it is downright impossible to extract meaningful data from experiments. This applies for example at low temperatures when thermodynamic equilibrium cannot be reached reliably.

For illustration consider the W–Ti system as analyzed using a combination of first-principles calculations, lattice models and Monte Carlo simulations. The analysis reveals a ground state structure at 80% tungsten concentration [19]. At low temperatures this configuration is much more likely to be observed than other configurations leading to a distinct feature in the mixing energy (Fig. 1.2a) and even more so the mixing entropy (Fig. 1.2b). These features are absent in CALPHAD assessments of the W–Ti system, which instead commonly assume complete immiscibility, i.e. a positive mixing energy throughout. This approximation is owed to the fact that W is a refractory metal, which renders experimental data below approximately 1300 K unreliable.

The combination of first-principles calculations (commonly density functional theory (DFT)) and lattice models (usually alloy cluster expansion (CEs)) allows one to accurately predict the behavior of multi-component mixtures with little or no experimental input. In the present thesis the application range of these techniques is further extended to analyze ordering in inorganic clathrates.

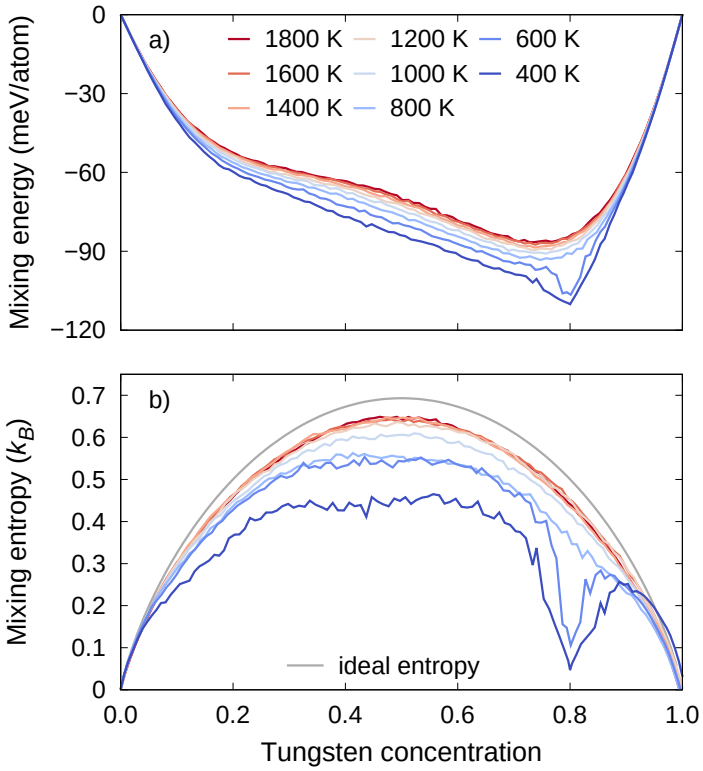

Figure 1.2: (a) Mixing energy and (b) mixing entropy as a function of composition.

1.3 Thesis guide

In this thesis a series of inorganic clathrates (chapter 2) has been investigated. Clathrates are inclusion compounds with complex ordering of the constituent atoms. For the systems of interest in this work, there are $\binom{46}{16} \approx 10^{12}$ possible ways to place the atoms in the unit cell, excluding symmetry. One therefore requires extremely efficient means to evaluate the energies of different configurations in order to investigate the vast configurational space of these systems. In this work this is achieved by means of cluster expansions (chapter 3) while Monte Carlo simulations are employed to obtain thermodynamical averages (chapter 4). The results of this work have been published in two peer-reviewed journal articles (chapter 5).

2

Inorganic clathrates

Inorganic clathrates constitute a class of inclusion compounds that exhibit a cage-like framework in which the cages are occupied by guest atoms or small molecules [20, 21]. The guest atoms, which are undersized relative to their respective cage, can act as so-called rattlers, lowering the lattice thermal conductivity. The framework structure can support a rather wide range of compositions, from binary systems to ternary and higher order. The availability of different compositions and the resulting variability of the distribution of elements in the framework provide opportunities for optimizing material properties. Inorganic clathrates have been studied in particular as potential high-performance thermoelectric materials due to their low intrinsic thermal conductivity [22, 23], suitable band gap [24, 25, 26], good dopability, and compositional variability [13].

2.1 Crystal structure

Inorganic clathrate can be categorized according to their symmetry [20, 21]. The present thesis focuses on type I clathrates, which have received the most attention so far (Fig. 2.1). The framework structure of type I inorganic clathrates contains 46 tetrahedrally coordinated host atoms in the unit cell. It is the geometrical arrangement of these 46 atoms that provides eight voids (or cages) per unit cell for the guest atoms. There are two smaller dodecahedral cages and six larger tetrakaidecahedral cages. The crystal structure belongs to the cubic space group $Pm\bar{3}n$. In Wyckoff notation the center of the cages are $2a$ and $6d$ for the dodecahedral and tetrakaidecahedral cages, respectively, whereas the framework atoms occupy Wyckoff sites $6c$, $16i$ and $24k$.

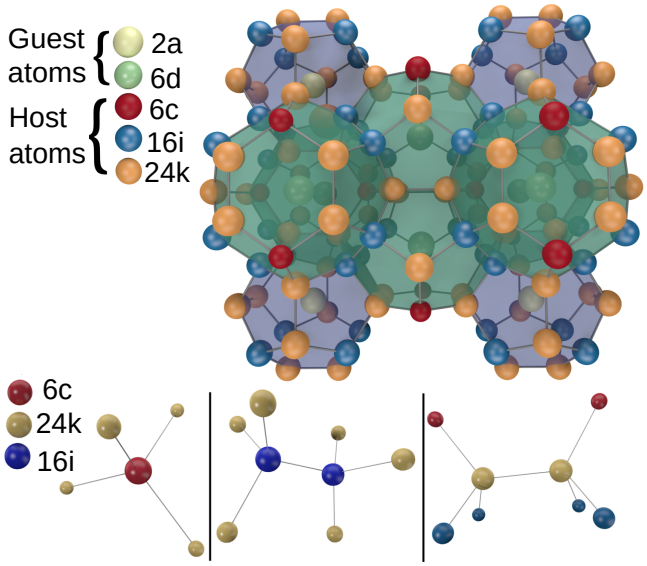

Figure 2.1: Crystal structure of type I clathrates. The guest species (Ba) occupies Wyckoff sites of type 2*a* and 6*d*, while the host species (Ga, Al, Ge, Si) occupy Wyckoff sites of type 6*c*, 16*i*, and 24*k*. The configurations in the bottom row illustrate the environments for 6*c*, 16*i*, and 24*k* sites, respectively.

2.2 The Zintl concept

The Zintl concept provides a rationale for the stoichiometry of semiconducting clathrates. It requires four electrons to be available for each tetrahedrally bonded host atom, while the guest atom is assumed to donate its valence electrons to the host framework. The number of electrons required to form bonds between the 46 host atoms is thus 184.

The general formula for type I clathrates is $A_8M_xM'_{46-x}$. The clathrates studied in this thesis are comprised of Ba, Ga/Al, Ge/Si for A, M and M' respectively. Each Ba atom can donate two electrons, whereas Ga/Al and Ge/Si provide three and four valence electrons, respectively. Therefore, for $x = 16$ all bonds saturated. Lowering the Ga/Al composition thus leads to electron deficiency and the material is expected to be n-doped. Likewise, increasing the Ge/Si concentration creates a p-doped material. As a thermoelectric element requires both an n-doped and p-doped material to function, clathrates can in principle achieve both of these limits by variation of the composition.

2.3 Ordering

The host framework of inorganic clathrates usually comprises several different Wyckoff sites (in the present work $6c$, $16i$, and $24k$), which occupied by several different species (here Ga, Ge, Si, Al). If the different sites were occupied statistically one would expect for a stoichiometric sample ($x = 16$) $16/46 \approx 35\%$ of the sites to be occupied by Al or Ga. Measurements of the so-called site occupancy factor (SOFs) reveal, however, dramatic deviations from this average [13]. The deviation from an entropically desirable random occupation (i.e., 35%) indicates that the interaction between the constituents plays a crucial role and the experimentally observed ordering (i.e. SOFs) arise from a competition of energy and entropy.

The SOFs vary between the compounds and can also show strong, non-monotonic variations with the stoichiometry [13, 14, 15]. Furthermore, the SOFs show an impact on transport properties [16]. Hence, understanding the ordering of these materials is crucial for understanding their thermoelectric performance.

2.3.1 Empirical rules for SOFs

A set of guideline rules for the SOFs has been formulated on the basis of a range of experimental data [13]. They are mostly based on the environment of each Wyckoff site and the observation that direct bonds between trivalent atoms species are unfavorable. There are three different bonding environments, one for each Wyckoff site. The $6c$ sites have four $24k$ sites as nearest neighbors; the $24k$ sites have one $6c$ site, two $16i$ sites and one $24k$ site as nearest neighbors; finally, the $16i$ sites have three $24k$ site and one $16i$ site as nearest neighbors.

The geometry of the lattice thus leads to the following set of rules.

1. A $6c$ site has no other $6c$ sites in its surrounding and hence $6c < 100\%$.
2. The same argument can be made to the other two sites giving $24k < 50\%$ and $16i < 50\%$ since there is one $24k - 24k$ and one $16i - 16i$ pair per $24k$ site and $16i$, respectively.
3. Furthermore, $6c + 24k < 100\%$ since the $6c$ site binds to four $24k$ sites and the sum of the SOFs should be below 100%.
4. Also $16i + 24k \leq 50\%$. As pointed out in Ref. [14], however, this rule is too restricted and the condition to avoid trivalent nearest neighbors is actually $16i + 24k \leq 83.3\%$.

The violation of the last rule can also be seen in papers I and II in this thesis, where we present ground state SOFs with no trivalent nearest neighbor and $16i + 24k = 56.25\%$. With that small modification, the rules, based on simple assumptions, are in good agreement with the experimental data.

2.3.2 Guideline for variations of SOFs

The rules described above cannot provide a direct rationale for explaining the *variations* in the SOFs as a function of composition. Different components in the framework structure will show different variations due to composition [15]. Here, atomic scale simulation, e.g., based on cluster expansions parametrized using first-principles calculations, provide a way to obtain very detailed information [16, 27, 15]. In Ref. [15] it was found that Al-based clathrates ($Ba_8Al_xGe_{46-x}$, $Ba_8Al_xSi_{46-x}$) exhibit a strong, non-monotonic variation of the SOFs with Al content, whereas Ga-based clathrates ($Ba_8Ga_xGe_{46-x}$, $Ba_8Ga_xSi_{46-x}$) show a monotonic dependence on Ga content. It was argued that the differences observed were largely due to the Al–Al repulsion being twice as strong as the Ga–Ga repulsion. In other words, the Al-based clathrates demonstrate a more extreme SOF behavior in order to avoid costly Al–Al bonds. The observations and conclusions from paper [15] have not been transferred to a general SOF model of clathrates. The results, however, demonstrate that information from first-principles calculations provides an efficient and accurate means to predict and rationalize ordering in these materials.

3

Alloy cluster expansions

3.1 Introduction

The partition function $\mathcal{Z}$, contains all thermodynamic information of a system To compute $\mathcal{Z}$ one needs to calculate the potential energy for each possible microstate of the system. For this to be feasible, however, one needs very efficient energy calculations. To this end, the alloy CE technique provides a computationally efficient and accurate way of calculating the energy for different microstates. In the cluster expansion formalism the system being operated on is described by the occupation vector $\boldsymbol{\sigma}$ where σ_i can, for a binary system, assume a value of either 0 or 1 depending on if an A or B atom is on lattice point i (Fig. 3.1). Even though the cluster expansion acts on a perfect lattice it can still capture the contribution of relaxations of the atoms by mapping relaxed structures onto the perfect lattice.

The remainder of the chapter is organized as follows:

- The definition of a cluster can be found in section 3.2
- The formal theory of cluster expansions can be found in section 3.3.
- The construction of a cluster expansion and how it can be seen as solving a linear equation system can be found in section 3.4
- In section 3.5 the concept of compressive sensing is introduced. compressive sensing (CS) are a class of algorithms that can be used to find the effective cluster interaction (ECIs) that describe a CE by solving the linear equation system.

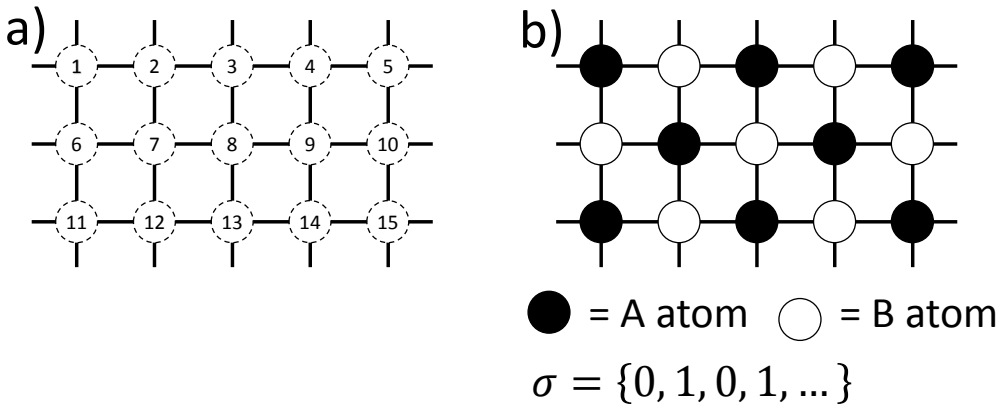

Figure 3.1: a) A fixed lattice with numbered lattice points. b) When the lattice points are occupied by atoms the state of the system can be described by the $\boldsymbol{\sigma}$ vector where σ_i is the occupation on site i.

- Cross-validation is introduced in section 3.6. Cross-validation is used both to estimate the error of the obtained ECIs and can also be used to find the best ECIs.

- An example of the procedure for constructing a cluster expansion for a simple binary system can be found in section 3.7

3.2 Definition of a cluster

A cluster is defined as a set of lattice points, $\boldsymbol{\alpha} = \{\sigma_1, \sigma_2, ..., \sigma_n\}$. A cluster is thus associated with a cell and possible periodic boundary conditions. The order of a cluster is defined as the number of lattice points in the cluster. A cluster of order 1 is called a singlet and order 2, 3 and 4 are called pair, triplet and quadruplet respectively. The radius, or the size, of the cluster is defined as the average distance of all the lattice points from the geometric center of the cluster. For a given lattice a set of clusters can conveniently be defined as a vector of cutoffs, $\boldsymbol{r}^{\text{cutoff}}$. The set of clusters will contain all clusters of order $i + 2$ with a maximum interatomic distance of less then or equal to r_i^{cutoff}. Figure 3.3 shows the clusters with the smallest radius of a body-centered cubic (bcc) lattice up to sixth order.

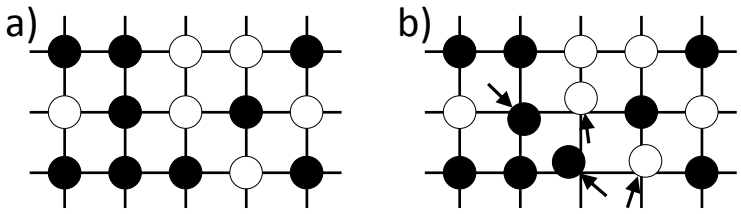

Figure 3.2: Two microstates of the lattice with the same $\boldsymbol{\sigma}$ vector. a) The atoms sit perfectly on the underlying lattice. b) Some atoms have small displacements away from the ideal positions.

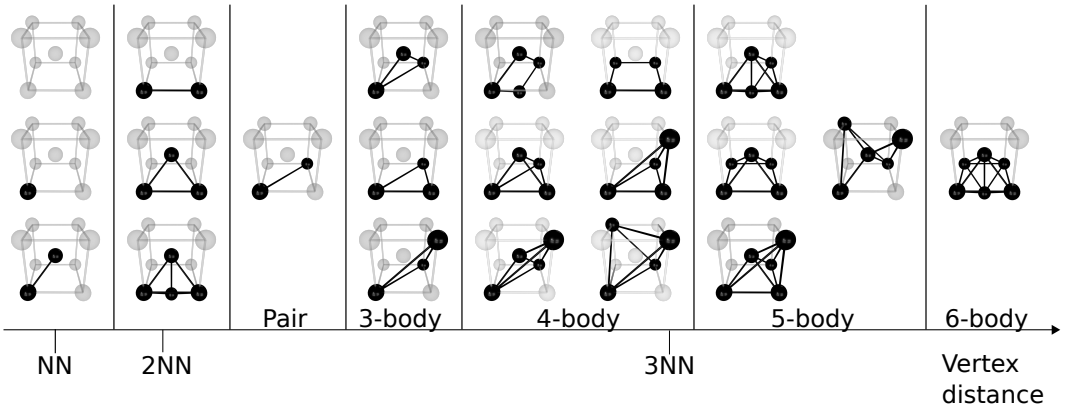

Figure 3.3: Illustration of the smallest clusters up to sixth order in a bcc lattice.

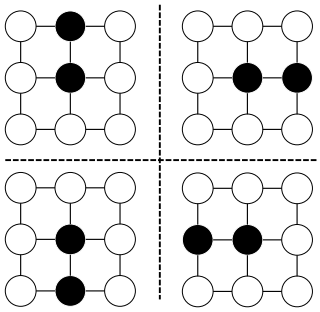

Figure 3.4: Symmetrically equivalent first nearest neighbor pair clusters in a square lattice. These can be found by taking any one of the specific decorations and repeatedly applying a 90 degree rotation until all four equivalent clusters are found.

3.3 Formal theory

A CE is able to represent *any* function of the configuration, $f(\boldsymbol{\sigma})$ if one can construct a complete orthogonal basis of functions with respect to the scalar product [28]

$$\langle f, g \rangle = \frac{1}{M^N} \sum_{\boldsymbol{\sigma}_1 ..., \boldsymbol{\sigma}_{M^N}} f(\sigma) g(\sigma)) \tag{3.1}$$

where $f(\boldsymbol{\sigma})$ and $g(\boldsymbol{\sigma})$ are two arbitrary functions of the configuration, M is the allowed number of elements and N is the number of lattice points in $\boldsymbol{\sigma}$.

3.3.1 Point functions

For each lattice point p we define the M orthogonal point functions $\Theta_n(\sigma_p)$

$$\Theta_n(\sigma_p) = \begin{cases} 1 & \text{if } n = 0 \\ -\cos\left(\pi(n+1)\sigma_p/M\right) & \text{if } n \text{ is odd} \\ -\sin\left(\pi n \sigma_p/M\right) & \text{if } n \text{ is even.} \end{cases} \tag{3.2}$$

It can be verified that these point functions form an orthogonal set over all possible occupation numbers [29],

$$\langle \Theta_n, \Theta_{n'} \rangle = \sum_{\sigma_p=0}^{M-1} \Theta_n(\sigma_p)\Theta_{n'}(\sigma_p) = \begin{cases} 0 & \text{if } n \neq n' \\ \neq 0 & \text{if } n = n'. \end{cases} \tag{3.3}$$

For example, in a three component system ($M = 3$) with $\sigma_p = \{0, 1, 2\}$ the possible point functions are

$$\Theta_0(\sigma_i) = 1, \quad \Theta_1(\sigma_i) = -\cos 2\pi \frac{\sigma_i}{3}, \quad \Theta_2(\sigma_i) = -\sin 2\pi \frac{\sigma_i}{3}.. \tag{3.4}$$

3.3.2 Orthogonal basis

With these point functions an orthogonal set of functions $\Pi_\alpha^{(s)}(\boldsymbol{\sigma})$ in the space of the M^N configurations on the lattice can be produced by generating the point functions for all possible combinations of s and lattice points $\boldsymbol{\alpha}$. So for a cluster of lattice sites $\boldsymbol{\alpha} = \{1, 2, ..., |\boldsymbol{\alpha}|\}$, and a vector of allowed point function indices, $s = n_1, n_2, ..., n_l$ the basis functions are given by,

$$\Pi_\alpha^{(s)}(\sigma) = \Theta_{n_1}(\sigma_1)\Theta_{n_2}(\sigma_2)...\Theta_{n_l}(\sigma_\alpha), \tag{3.5}$$

and it can be verified that these form an orthogonal set[29],

$$\left\langle \Pi_\alpha^{(s)}, \Pi_\beta^{(s')} \right\rangle = \delta_{\alpha\beta}\delta_{ss'}. \tag{3.6}$$

Since the basis functions $\Pi_\alpha^{(s)}$ form an orthogonal set we can express any function of the configuration as

$$f(\sigma) = \sum_\alpha \sum_s f_{\alpha s} \Pi_\alpha^{(s)}(\sigma). \tag{3.7}$$

Since all basis functions $\Pi_\alpha^{(s)}$ have one configuration invariant component that is equal to 1 when $s = \{0, 0, ..., 0\}$ we can exclude this term from the sum in Eq. (3.7) to obtain

$$f(\sigma) = f_0 + \sum_\alpha \sum_s f_{\alpha s} \Pi_\alpha^{(s)}(\sigma). \tag{3.8}$$

Finally, for practical reasons, we modify Eq. (3.8) by averaging over each distinct cluster and point functions and multiply with the multiplicity and arrive at the final expression for our cluster expansion function

$$f(\sigma) = f_0 + \sum_\alpha \sum_s \left\langle \Pi_\alpha^{(s)}(\sigma) \right\rangle_{\alpha'} m_\alpha^{(s)} J_\alpha^{(s)}. \tag{3.9}$$

Here, the summation is carried out over all symmetrically distinct clusters of lattice points. The $\langle ... \rangle_{\alpha'}$ function takes the average over the basis functions for all clusters α' that are symmetry equivalent to α. $J_\alpha^{(s)}$ are the ECIs, which determine a specific cluster expansion. Finally, $m_\alpha^{(s)}$ is the multiplicity of cluster α for a specific combinations of point functions $\boldsymbol{s}$.

3.3.3 Further considerations

As we have seen from the construction of the basis, all combinations of point functions are required for constructing the basis. For a binary system where only the first point function was needed the permutations of these point functions for any cluster order are all equal. For a ternary system both the first and second point functions are needed and thus a pair will have four different combinations of the point functions, i.e. (1, 1), (1, 2), (2, 1) and (2, 2). As will be shown now, not all of these permutations will result in additional parameters in the CE due to symmetry. If our pair cluster $\alpha = \{\sigma_1, \sigma_2\}$ can be transformed under periodic boundary conditions and the lattice translational and rotational symmetry operations to construct the symmetrically equivalent cluster $\alpha' = \{\sigma_2, \sigma_1\}$ then the choice of the ordering in $\alpha = \{\sigma_1, \sigma_2\}$ must produce the same contribution to Eq. (3.9) as choosing $\alpha = \{\sigma_2, \sigma_1\}$. Writing out the different choices we have

$$\begin{aligned} \Pi_\alpha^{1,1}(\sigma) &= \Theta_1(\sigma_1)\Theta_1(\sigma_2) \\ \Pi_\alpha^{1,2}(\sigma) &= \Theta_1(\sigma_1)\Theta_2(\sigma_2) \\ \Pi_\alpha^{2,1}(\sigma) &= \Theta_2(\sigma_1)\Theta_1(\sigma_2) \\ \Pi_\alpha^{2,2}(\sigma) &= \Theta_2(\sigma_1)\Theta_2(\sigma_2) \end{aligned} \tag{3.10}$$

and for α' we get

$$\begin{aligned}\Pi_{\alpha}^{1,1}(\sigma) &= \Theta_1(\sigma_2)\Theta_1(\sigma_1)\\ \Pi_{\alpha}^{1,2}(\sigma) &= \Theta_1(\sigma_2)\Theta_2(\sigma_1)\\ \Pi_{\alpha}^{2,1}(\sigma) &= \Theta_2(\sigma_2)\Theta_1(\sigma_1)\\ \Pi_{\alpha}^{2,2}(\sigma) &= \Theta_2(\sigma_2)\Theta_2(\sigma_1).\end{aligned} \tag{3.11}$$

The point functions basis functions (1, 1) and (2, 2) are thus symmetric in terms of permuting the lattice points. For (1, 2) and (2, 1) there is, however, an asymmetry and the choice of the order in the cluster α will matter for the end result in Eq. (3.9). The choice of ordering the lattice points in a cluster is completely arbitrary and should not matter to the final result of the cluster expansion. To circumvent the choice of ordering of lattice points in a cluster one instead only use the point functions (1, 1), (1, 2) and (2, 2) for this cluster. Additionally, for the cluster basis (1, 2) one uses both possibilities of ordering the pair. The final basis functions for cluster α then become the following

$$\begin{aligned}\Pi_{\alpha}^{1,1}(\sigma) &= \Theta_1(\sigma_1)\Theta_1(\sigma_2)\\ \Pi_{\alpha}^{1,2}(\sigma) &= \Theta_1(\sigma_1)\Theta_2(\sigma_2) + \Theta_1(\sigma_2)\Theta_2(\sigma_1),\\ \Pi_{\alpha}^{2,2}(\sigma) &= \Theta_2(\sigma_1)\Theta_2(\sigma_2)\end{aligned} \tag{3.12}$$

where it it is apparent that the choice of the order in the cluster has no effect on its representation in the cluster space. Note that in Eq. (3.9) the multiplicity $m_{\alpha}^{(1,2)}$ will be twice as large as the other multiplicities.

3.3.4 Symmetrically indistinct clusters

This section describes the identification of equivalent clusters by using symmetry operations. For the n-body cluster $\alpha = \{\sigma_0, \sigma_1, .., \sigma_{n-1}\}$ the symmetrically equivalent clusters are found by converting the lattice points into fractional positions $\alpha = \{\boldsymbol{r}_0, \boldsymbol{r}_1, .., \boldsymbol{r}_{n-1}\}$. A symmetry operation $\hat{\boldsymbol{s}}$ consists of a linear transformation by a 3×3 matrix,$\bar{\gamma}$ and a translation $\boldsymbol{\tau}$

$$\boldsymbol{r}' = \hat{\boldsymbol{s}}\boldsymbol{r} = \bar{\gamma}\boldsymbol{r} + \boldsymbol{\tau}. \tag{3.13}$$

A lattice typically have a number of associated symmetry operations $\boldsymbol{S} = \hat{\boldsymbol{s}}_0$, $\hat{\boldsymbol{s}}_1$, ..., $\hat{\boldsymbol{s}}_{m-1}$. These symmetry operations can then be used to produce m symmetrically indistinct clusters where the cluster i is given by

$$\alpha_i = \{\hat{\boldsymbol{s}}_i\boldsymbol{r}_0, \hat{\boldsymbol{s}}_i\boldsymbol{r}_1, .., \hat{\boldsymbol{s}}_i\boldsymbol{r}_{n-1}\}. \tag{3.14}$$

Additionally, depending on the periodic boundary conditions, one can also translate all positions in a cluster with multiples of the unit cell vectors. Figure 3.4 illustrates the set of equivalent clusters for the nearest neighbor pair.

3.4 Construction of a cluster expansion

Equation (3.9) can represent any function of the configuration $f(\boldsymbol{\sigma})$. The task remaining is in the construction of a cluster expansion is to find the ECIs. To this end, one requires reference data in the form of a set of configurations $\{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, ..., \boldsymbol{\sigma_n}\}$ as well as target data $\{E_1, E_2, ..., E_n\}$. The sums in Eq. (3.9) can be replaced with a dot product

$$f(\boldsymbol{\sigma}) = f_0 + \sum_{\alpha} \sum_{s} \left\langle \Pi_{\alpha}^{(s)}(\sigma) \right\rangle_{\alpha'} m_{\alpha}^{(s)} J_{\alpha}^{(s)} = \boldsymbol{\omega}(\boldsymbol{\sigma}) \cdot \boldsymbol{J}, \tag{3.15}$$

where

$$\boldsymbol{\omega}(\boldsymbol{\sigma}) = \Bigg\{ 1, \left\langle \Pi_{\alpha_1}^{(s_\alpha)}(\sigma) \right\rangle_{\alpha_1'} m_{\alpha_1}^{(s_{\alpha_1})}, ... \left\langle \Pi_{\alpha_1}^{(s'_{\alpha_1})}(\sigma) \right\rangle_{\alpha_1'} m_{\alpha_1}^{(s'_{\alpha_1})}, \\ ... \left\langle \Pi_{\alpha_n}^{(s_{\alpha_n})}(\sigma) \right\rangle_{\alpha_n'} m_{\alpha_n}^{(s_{\alpha_n})} \Bigg\},$$

and $\boldsymbol{J}$ denotes the vector of ECIs where $J_0 = f_0$. The vector $\boldsymbol{\omega}(\boldsymbol{\sigma})$ is commonly called cluster vector. Note that it can sometimes be useful to exclude $m_{\alpha}^{(s)}$ from $\boldsymbol{\omega}$ and let the target values E_i refer to the primitive unit cell. This will ensure all elements in $\boldsymbol{\omega}$ are in the interval $[-1, 1]$ and avoid a bias due to the number of elements in $\boldsymbol{\sigma}$. Now we can cast the problem of finding the ECIs in the form of a linear equation

$$\begin{bmatrix} \boldsymbol{\omega}(\boldsymbol{\sigma}_1) \\ \boldsymbol{\omega}(\boldsymbol{\sigma}_2) \\ \vdots \\ \boldsymbol{\omega}(\boldsymbol{\sigma_n}) \end{bmatrix} \begin{bmatrix} J_{\alpha_1}^{(s)} \\ J_{\alpha_2}^{(s)} \\ \vdots \\ J_{\alpha_n}^{(s)} \end{bmatrix} = \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_n \end{bmatrix} \tag{3.16}$$

The effective cluster interactions, i.e the $J_{\alpha}^{(s)}$ of each cluster $\boldsymbol{\alpha}$, are unknown and since the number of clusters is in principle infinite there is an infinite number of unknown parameters to determine. Based on physical intuition we, however, expect that physical interactions are short-ranged and few-bodied. Therefore, if we construct our basis functions starting from singlets and geometrically small pairs, triplets etc. the CE is expected to converge quickly to yield an acceptable tolerance and the number of unknown parameters remains manageable.

There have also been advances in the algorithms that solve Eq. (3.16). These CS algorithms can find the ECIs even if the problem is severely under-determined, i.e. the number of unknowns are much larger than the number of available data points. The choice of when to truncate is thus not so important which effectively removes much of the "human factor" when it comes to constructing a cluster expansion. Solving under determined problems and how to do validation is discussed in the following sections.

3.5 Compressive sensing

One way to solve Eq. (3.16) when the problem is underdetermined is by taking advantage of CS algorithms. The CS technique provides a simple and efficient way to extract the important ECIs and compute their values in one shot [30, 31, 32]. To demonstrate CS it helps to first define the l_p norm of a vector u

$$\|u\|_p = \left(\sum_i |u_i|^p \right)^{1/p}. \tag{3.17}$$

In the CS method the problem is solved by searching for the solution with the smallest l_1 norm while still reproducing the results with a given accuracy

$$J_{\mathrm{CS}} = \arg\min_J \left\{ \|\boldsymbol{J}\|_1 : \left\|\bar{\Pi}\boldsymbol{J} - \boldsymbol{E}\right\|_2 < \epsilon \right\}. \tag{3.18}$$

Where, in the case of cluster expansions, $\bar{\Pi}$ is the matrix of cluster vectors, $\boldsymbol{J}$ is the ECIs and $\boldsymbol{E}$ is the target properties. The form in Eq. (3.18) is inconvenient to work with and it is common practice to work with an unconstrained approach that minimizes the l_1 norm and the least squares sum of the fitting error

$$\boldsymbol{J} = \arg\min_J \left\{ \mu \|\boldsymbol{J}\|_1 + \frac{1}{2} \left\|\boldsymbol{E} - \bar{\Pi}\boldsymbol{J}\right\|^2 \right\}, \tag{3.19}$$

where the parameter μ controls the accuracy of the fit. A high value of μ leads to sparse solution but larger prediction error and vice versa. It is, however, difficult to efficiently implement mixed l_1 and l_2 minimization problems such as Eq.(3.19).

3.5.1 Split Bregman algorithm

Goldstein and Osher proposed the split Bregman algorithm which eliminates this problem [33]. The split Bregman iteration splits the l_1 norm of the solution from the objective function and replaces it with a variable $\boldsymbol{d}$ which then converges towards the l_1 term $\lim_{k\to\infty}(\boldsymbol{d}-\mu\boldsymbol{J}) = \boldsymbol{0}$, where k is the number of split Bregman iterations. To this end a least-squares l_2 term is added to the objective function to ensure that $\boldsymbol{d} = \mu\boldsymbol{J}$

$$\boldsymbol{J} = \arg\min_{J,d} \left\{ \|\boldsymbol{d}\|_1 + \frac{1}{2} \left\|\bar{\Pi}\boldsymbol{J} - \boldsymbol{E}\right\|^2 + \frac{\lambda}{2} \|\boldsymbol{d} - \mu\boldsymbol{J}\|^2 \right\}. \tag{3.20}$$

This formulation is advantageous because the minimization involving the quadratic form $\frac{1}{2}\left\|\bar{\Pi}\boldsymbol{J} - \boldsymbol{E}\right\|^2$ does not involve any l_1 terms and can be minimized efficiently using efficient l_2 minimization algorithms. The split Bregman algorithm comprises the following steps

$$\boldsymbol{J}^{k+1} = \arg\min_J \left\{ \frac{1}{2} \left\|\bar{\Pi}\boldsymbol{J} - \boldsymbol{E}\right\|^2 + \frac{\lambda}{2} \left\|\boldsymbol{d}^{\,k} - \mu\boldsymbol{J} - \boldsymbol{b}^k\right\|^2 \right\} \tag{3.21}$$

$$\boldsymbol{d}^{k+1} = \arg\min_{d} \left\{ ||\boldsymbol{d}||_1 + \frac{\lambda}{2} \left\| \boldsymbol{d} - \mu \boldsymbol{J}^{k+1} - \boldsymbol{b}^k \right\|^2 \right\} \tag{3.22}$$

$$\boldsymbol{b}^{k+1} = \boldsymbol{b}^k + \mu \boldsymbol{J}^{k+1} - \boldsymbol{d}^{k+1} \tag{3.23}$$

Where in Eq.(3.23) the residual after iteration k is added back to the residual vector $\boldsymbol{b}^{k+1}$ for the next iteration in style with a Bregman iteration which result in a quicker convergence [34]. Starting from $\boldsymbol{d}^0 = 0$, $\boldsymbol{b}^0 = 0$ and $\boldsymbol{J}^0 = 0$. First, the l_2 minimization in Eq.(3.21) is solved, the second step, Eq.(3.22) separates into individual vector components and is solved by shrinkage,

$$\boldsymbol{d}_n^{k+1} = \text{shrink}(\mu \boldsymbol{J}_n^{k+1} + \boldsymbol{b}_n^k, 1/\lambda) \tag{3.24}$$

which is defined by

$$\text{shrink}(y, \alpha) \equiv \text{sign}(y)\max(|y| - \alpha, 0) \tag{3.25}$$

Shrinkage decreases the absolute magnitude of the y vector by α and sets it to zero if $y \leq \alpha$. This procedure is then usually repeated until the J vector has converged within a target tolerance.

3.6 Cross-validation

When training a model such as a CE one requires a set of data points that can be used to train and test the obtained model. Commonly the data points is split up in a training set and a test set. The training set is the data points used as input to optimization algorithms that solve problems similar to equation (3.18). A training set has an associated training error defined as $\left\| \bar{\Pi} \boldsymbol{J} - \boldsymbol{E} \right\|_2$ where $\boldsymbol{J}$ is the parameters obtained from training with the training set, $\boldsymbol{E}$ and $\bar{\Pi}$ is the target property and the description matrix for the training set respectively. The test set is data points which were not used in the training set. The test error is the l_2 norm of the prediction error of the test set. The usefulness of these different sets and errors come in when trying to find a good model. For example, if for a particular value of μ, the training error is zero but the test error is high, the trained model is excellent at predicting already seen data but are unable to predict new and unseen data. This is called overfitting and must be avoided. Another type of overfitting is when one instead find values of μ that finds the minimum test error. Even though the test set is not part of the training set, information about the training set still seeps into the training procedure since the parameters obtained in the training are optimal for the specific test set. Cross-validation (CV) is a way to overcome both these types of overfitting both when estimating the error and when finding a model. In CV the training and tests sets do not remain fixed which reduces the risk of overfitting to a specific training or testing set. The use of CV scores is widely accepted as the quantity for determining the accuracy of the CE.

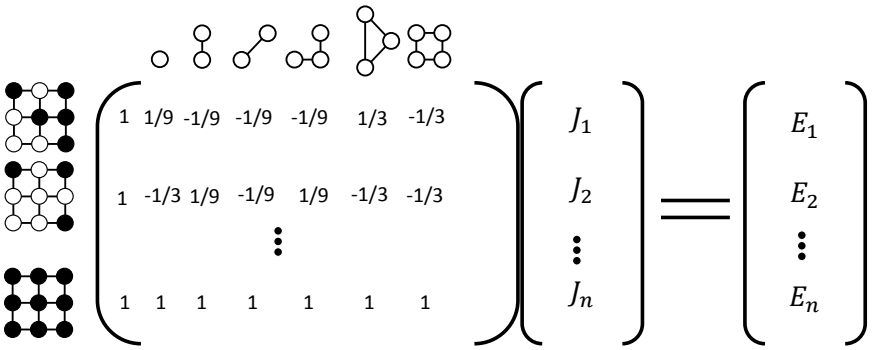

Figure 3.5: Cluster expansions can be constructed by solving a set of linear equation systems.

3.6.1 Leave-one-out cross validation

One type of cross validation is the leave-one-out CV (LOO-CV), which is defined as

$$(\mathrm{CV})^2 = \frac{1}{N}\sum_{n=1}^{N}(\hat{E}_{(n)} - E_n)^2, \tag{3.26}$$

where E_n is the calculated energy for structure n and $\hat{E}_{(n)}$ is the predicted value of the energy of structure n as calculated with the CE fit with the $(N-1)$ other structures. This method requires making N cluster expansions so it can be computationally expensive. When the training curve is steep and the number of available structures are few the LOO-CV can be a good estimator compared to other estimators which splits the available structures more and hence over estimate the error.

3.6.2 k-fold cross validation

Another commonly used estimator for the error are k-folds cross validations. Here, the data is randomly divided up into k evenly sized subsets. One of the k subsets will be left out for validation and the remaining k-1 subsets will be used for training. This is repeated for all k subsets and the final cross validation score is the average validation for the k validation scores. Note that k-fold validation reduces to LOO-CV when k is the number of available data points.

3.7 Cluster expansion for a binary system

To demonstrate the formalism developed in this chapter we will now explicitly describe the construction of a CE for simple binary system. In a binary system only one point function will be used in the cluster functions, $\Theta_1(\sigma_p) = -\cos\left(\pi\sigma_p\right)$. Only two occupation numbers are needed as well, 0 and 1, which have a corresponding point function value of -1 and $+1$ respectively. Hence a binary cluster expansions share a lot of similarities with the Ising model. Next, make the choice that white and black atoms are to be indicated by occupation 0 and 1 respectively. The average over symmetrically distinct clusters α' in equation (3.9) will for the singlet cluster be

$$\left\langle \Pi_{\alpha=\text{singlet}}(\boldsymbol{\sigma}) \right\rangle_{\alpha'} = \frac{N_{\boldsymbol{\sigma}_{\alpha'}=(1)} - N_{\boldsymbol{\sigma}_{\alpha'}=(0)}}{N_{\boldsymbol{\sigma}_{\alpha'}=(0\,\|\,1)}}. \tag{3.27}$$

For a pair the averages will become

$$\left\langle \Pi_{\alpha=\text{pair}}(\boldsymbol{\sigma}) \right\rangle_{\alpha'} = \frac{N_{\boldsymbol{\sigma}_{\alpha'}=(0,0)} + N_{\boldsymbol{\sigma}_{\alpha'}=(1,1)} - N_{\boldsymbol{\sigma}_{\alpha'}=(0,1)}}{N_{\boldsymbol{\sigma}_{\alpha'}=(0,0\,\|\,0,1\,\|\,1,1)}}, \tag{3.28}$$

where $N_{\boldsymbol{\sigma}_{\alpha'}=(i,j)}$ is the number of equivalent clusters that have occupation (i,j) or (j,i). To calculate values such as $N_{\alpha'=(0,0)}$ a summation is done over all index pairs (i,j) that are symmetrically equivalent. Hence, the construction of the cluster vector, $\boldsymbol{\omega}(\boldsymbol{\sigma})$, is obtained by summation and averaging of the cluster functions. All the $\boldsymbol{\alpha}$ can be precomputed so a summation can be performed very efficiently with a computer program. Then all available configurations can be mapped to cluster vectors, $\boldsymbol{\omega}$, which can constitute different training and test sets (Fig. 3.5) as described in the previous sections. Once a suitable value for the fitting parameter μ have been found, by using various cross validation techniques, the final set of ECIs, $\boldsymbol{J}_{\text{final}}$, can be obtained. The construction of the cluster expansion is now complete and the expanded property can for a configuration $\boldsymbol{\sigma}$ be computed very efficiently with $\boldsymbol{\omega}(\boldsymbol{\sigma})\boldsymbol{J}_{\text{final}}$.

4

Monte Carlo simulations

4.1 Monte Carlo integration

Monte Carlo (MC) methods represent a broad class of computer algorithms that are based on the use of random numbers to sample high-dimensional functions. MC integration is one example of such a technique where the average value of an integral

$$I = \int_0^1 dx f(x) = \langle f(x) \rangle \tag{4.1}$$

can be approximated by evaluating $f(x)$ at N points x_i chosen at random with uniform probability over the interval $[0, 1]$. The mean value becomes

$$I_N = \langle f \rangle = \frac{1}{N} \sum_{i=1}^{N} f(x_i) = \frac{1}{N} \sum_{i=1}^{N} f_i, \tag{4.2}$$

and the variance

$$\sigma_f^2 = \langle f^2 \rangle - \langle f \rangle^2 . \tag{4.3}$$

Such that the integral is approximated by

$$I = I_N \pm \frac{\sigma_f}{\sqrt{N}}. \tag{4.4}$$

By increasing the number of points N the error of the approximation becomes smaller and as $N \to \infty$ one approaches the correct value of I.

4.2 Importance sampling

An alternative approach to decreasing the error is to choose the points x_i based on the magnitude of $f(x)$. Consider a probability density function $p(x) > 0$ that is positive and normalized to 1 on $[0, 1]$. We can rewrite equation (4.1) to obtain

$$I = \int_0^1 dx f(x) = \int_0^1 dx \frac{f(x)}{p(x)} p(x) = \int_0^1 dx g(x) = \langle g(x) \rangle_p , \tag{4.5}$$

where $g(x) = f(x)/p(x)$ and the notation $\langle ... \rangle_p$ signifies that the average is obtained from sampling values of x according to $p(x)$. The mean value of the integral can then be written as

$$I_N = \langle f \rangle = \frac{1}{N} \sum_{i=1}^{N} g_i \tag{4.6}$$

and I can be approximated with

$$I = I_N \pm \frac{\sigma_g}{\sqrt{N}}. \tag{4.7}$$

Assume now that $p(x)$ follows the approximate behavior of $f(x)$, i.e. $p(x)$ has high probability density when $|f(x)|$ is large and vice versa. Then $g(x)$ will become a smoother function than $f(x)$ and $\sigma_g < \sigma_f$. This approach of choosing values of x where $|f(x)|$ is large is called importance sampling.

4.3 Thermodynamic integration on the lattice

Now, instead of approximating a one-dimensional integral let us consider a binary atomic lattice $\boldsymbol{\sigma}$ at a certain temperature T and find the average of some quantity A that depends on the configuration. Let the atomic lattice vector $\boldsymbol{\sigma}$ have dimensionality N, where N is the number of lattice points and the elements of the vector can assume values of either 0 or 1. The average value is then

$$\langle A \rangle = \frac{1}{M^N} \sum_{\text{all possible } \boldsymbol{\sigma}} A(\boldsymbol{\sigma}) P(\boldsymbol{\sigma}). \tag{4.8}$$

The probability of finding the system at $\boldsymbol{\sigma}$ is $P(\boldsymbol{\sigma})$, which is defined as

$$P(\boldsymbol{\sigma}) = \frac{\exp\left(-U(\boldsymbol{\sigma})/k_B T\right)}{\sum_{\text{all possible } \boldsymbol{\sigma}'} \exp\left(-U(\boldsymbol{\sigma}')/k_B T\right)}, \tag{4.9}$$

where $U(\boldsymbol{\sigma})$ is the internal energy of the configuration and k_B is the Boltzmann constant. Carrying out the sum in equation (4.8) is unfortunately impossible to

carry out exactly. For illustration, consider that for a system with 100 sites the number of possible configurations, ignoring symmetry, is $2^{100} \approx 10^{30}$. Also assume that each evaluation of $A(\boldsymbol{\sigma})$ requires only one floating point operation (FLOP). The time to calculate the sum by using the largest computer cluster available [1] would still require around 10^6 times longer than the age of the universe. This demonstrates the necessity to find efficient methods to solve to Eq. (4.8).

4.4 The Metropolis algorithm

A naive approach to approximate Eq. (4.8) using a MC approach is to generate a large number of different configurations, $\boldsymbol{\sigma}$, where each element is randomly assigned a value of either 0 or 1, and estimating the average. This approach would, however, lead to very slow convergence due the probability function $P(\boldsymbol{\sigma})$ commonly being a very sharp function, which would lead to $P(\boldsymbol{\sigma})$ being close to zero for most choices of $\boldsymbol{\sigma}$. A more prudent approach would be to try to apply the importance sampling approach introduced earlier. A direct implementation of importance sampling is difficult, however, since it is not obvious how to efficiently generate configurations $\boldsymbol{\sigma}$ according to a suitable probability. Instead, an efficient way to generate configurations is to implement a so-called Markov chain, where each new configuration generated is based on a probability ratio that depends on the previous configuration.

The first adaptation of such a approach was introduced in 1953 by Metropolis *et al.* to determine the equation of state for a hard sphere liquid [35]. It is based on the understanding that thermodynamic averaging only requires knowledge of relative rather than absolute probabilities such as in Eq. (4.8). It uses a Markov chain to generate configurations that are more important by rejecting configurations that are unlikely, similar to importance sampling.

4.4.1 Markov chain Monte Carlo

A Markov chain is a stochastic process, in which a system undergoes transitions from one state to another. The Markov process is characterized by a lack of memory of where it has been. The future of the chain depends solely on the current state. The transition probability is given by a transition matrix

$$T(X \to X') \equiv T_{XX'}, \tag{4.10}$$

for a transition from state X to state X'. The transition probability has to satisfy

$$0 \leq T_{XX'} \leq 1 \tag{4.11}$$

[1] The Sunway TaihuLight in China is currently the largest non-distributed computer cluster with a peak performance at about 125 PFLOPS.

and the probability has to be normalized

$$\sum_{X'} T_{XX'} = 1. \tag{4.12}$$

The task is to generate a Markov chain of configurations such that they have a distribution proportional to the Boltzmann factor and this distribution should be independent on the position of the chain and of the initial configuration. The Markov chain can exhibit these properties under certain conditions, at least for a sufficiently long time so that the configuration can loose memory of its initial state. These conditions are:

- The Markov chain needs to be irreducible, that is every configuration included in the ensemble should be accessible from every other configuration within a finite number of steps.
- There should be no periodicity. Periodicity means that it is not possible to revisit a configuration except after $t = nk$ steps, $n = 1, 2, 3 \ldots$, where k is fixed.

A Markov chain that satisfies these conditions is called ergodic. If the Markov chain is ergodic it converges to a unique stationary distribution. The transition probability needs to be chosen such that the stationary distribution is the desired distribution. To assure this, consider the stationary distribution $\rho(X)$; one can also introduce a new function $\rho(X, t)$, which gives the probability of finding configuration X after t Markov steps, which for an ergodic chain becomes independent of t if t is large. This function can change from one step to another by

- going from X at step t to X' at $t + 1$ leading to a decrease in $\rho(X)$
- going from X' at step t to X at $t + 1$ leading to an increase in $\rho(X)$

that can be summarized with

$$\rho(X, t+1) - \rho(X, t) = -\sum_{X'} T(X \to X')\rho(X, t) + \sum_{X'} T(X' \to X)\rho(X', t). \tag{4.13}$$

This equation is called the master equation. The stationary solution of this equation is found by requiring $\rho(X, t + 1) = \rho(X, t)$ so we have

$$\sum_{X'} T(X \to X')\rho(X, t) = \sum_{X'} T(X' \to X)\rho(X', t). \tag{4.14}$$

Leaving out the t-dependence, which is allowed due to the "memory loss" of a Markov chain, yields

$$T(X \to X')\rho(X) = T(X' \to X)\rho(X'), \tag{4.15}$$

which is known as the condition of detailed balance. This means that in equilibrium the average number of steps that results in the system leaving state X must be exactly equal to the number of steps from all other states X' to X. This means that $\rho(X)$ and $\rho(X')$ do not change. Since this is true for all pairs of X and X' the probability distributions will remain stationary. Reformulating the detailed balance condition with the transition probability in this form yields

$$T(X \to X') = \omega_{XX'} A_{XX'}, \tag{4.16}$$

where $\omega_{XX'}$ is the probability for going from state X to state X' and is symmetric $\omega_{XX'} = \omega_{X'X}$. $A_{XX'}$, which must lie between 0 and 1, is the acceptance probability for actually committing the change. The detailed balance condition can then be expressed as

$$\frac{A_{XX'}}{A_{X'X}} = \frac{\rho(X')}{\rho(X)}. \tag{4.17}$$

If the sought after distribution is the Boltzmann distribution, $\rho(X) = \exp -\beta U(X)$, there are a number of choices for the acceptance probability $A_{XX'}$ that will accomplish this. The choice of Metropolis *et al.* was

$$\begin{cases} A_{XX'} & = \exp \beta \left[U(X') - U(X)\right] & \text{if } \rho(X') < \rho(X) \\ A_{XX'} & = 1 & \text{if } \rho(X') \geq \rho(X). \end{cases} \tag{4.18}$$

The Metropolis algorithm can now be formulated as follows:

- Starting from a state X, make a small trial move into a new state X' with a probability of $\omega_{XX'}$.
- Compare the weights of the distribution for the different states $\rho(X)$ and $\rho(X')$. $A_{XX'}$, the acceptance probability, is chosen equal to 1 if $\rho(X') > \rho(X)$ else it is chosen to be equal to $\rho(X')/\rho(X)$.
- The new state X' is accepted with probability $A_{XX'}$ (the system moves from X to X') and is rejected with probability $1 - A_{XX'}$ (the system remains in state X). To decide if a state is accepted or not a random number is generated uniformly in the range $[0, 1]$ and compared to the acceptance probability. If the random number is larger than the acceptance probability the trial move is accepted.

Since each trial move is only a small change in the configuration there is an inherent correlation between the states X and X'. There is thus a correlation length s for the Markov chain and it is necessary to carry out s trial steps before reaching a new uncorrelated configuration. One MC step (or cycle) is defined as N trial steps, where N is the number of particles which is commonly used to approximate s. There is also a need to equilibrate the initial configuration meaning that it is necessary to run the Metropolis algorithm before the actual sampling commences.

4.5 Probability of a state

The sampling procedure described in the previous section requires the probability ratio of two microstates. Let us consider again the rigid binary lattice introduced above. Each possible combination of zeros and ones on the $\boldsymbol{\sigma}$ vector describes a microstate of the system. Consider the case where this system is in thermal contact with a heat reservoir with constant temperature and an infinitely large heat capacity, i.e. we can add and remove as much energy as necessary without affecting the temperature of the reservoir. Since the composition of the system might change there can also be an exchange of atoms between system and reservoir.

For an isolated system, all possible microstates are equally probable. The system of N sites considered here is, however, not isolated since it is in contact with a heat reservoir. Rather the *joint* system and the heat reservoir form an isolated system. The joint microstates of system and heat reservoir will therefore be equally probable.

Consider two microstates s_1 and s_2 with a corresponding number of accessible states $\Omega_R(s_1)$ and $\Omega_R(s_2)$ for each respective reservoir. At this point there is no way of telling what the actual probabilities of the different states are since the number of accessible states associated with the heat reservoirs is unknown. All we know at this point is that the probability of a state is proportional to the total number of accessible microstates, i.e. $P(s_1) \propto \Omega_R(s_1)$. Consider now the ratio of probabilities

$$\frac{P(s_1)}{P(s_2)} = \frac{\Omega_R(s_1)}{\Omega_R(s_2)}. \tag{4.19}$$

Rewriting this equation by using the definition of entropy $S = k_B \ln \Omega$ one obtains

$$\frac{P(s_1)}{P(s_2)} = \frac{e^{S_R(s_1)/k_B}}{e^{S_R(s_2)/k_B}} = e^{[S_R(s_1) - S_R(s_2)]/k_B}. \tag{4.20}$$

Now the ratio of probabilities depends on the change of entropy in the reservoir when going from state s_1 to state s_2. The change in the entropy of the reservoirs should be small since the system is small compared to the reservoir. Then we can use the thermodynamic identity

$$dS = \frac{1}{T}\left(\mathrm{d}U + P\mathrm{d}V - \mu \mathrm{d}N\right). \tag{4.21}$$

Since all ensembles discussed in this thesis have constant volume V and fixed number of total atoms (or sites) N, the expression can be simplified to $dS = S(s_2) - S(s_1) = \frac{1}{T}\left(\mathrm{d}U - \Delta\mu\Delta N_A\right) = -\left(E(s_2) - E(s_1) - \Delta\mu(N_A(s_2) - N_A(s_1))\right)$, where E is the internal energy of the system, $N_A(s_2)$ and $N_A(s_1)$ are the numbers of A atoms in system s_2 and s_1, respectively, and $\Delta\mu$ is the chemical potential difference

between species A and B

$$\frac{P(s_1)}{P(s_2)} = e^{(E(s_2)-E(s_1)-\Delta\mu\Delta N_A)/k_BT}. \tag{4.22}$$

As was shown in section 4.5 this ratio of probabilities is all that is needed to carry out a Metropolis MC simulation.

For the sake of completeness, however, let us determine the actual probability of the microstate. First we separate all terms in Eq. (4.22) related to s_1 to one side and terms related to s_2 to the other side,

$$P(s_1)e^{(E(s_1)-\Delta\mu N_A(s_1))/k_BT} = P(s_2)e^{(E(s_2)-\Delta\mu N_A(s_2))/k_BT}. \tag{4.23}$$

The right hand side does not depend on state s_1 and vice versa and must therefore be equal to a constant

$$P(s_1)e^{(E(s_1)-\Delta\mu N_A(s_1))/k_BT} = A \tag{4.24}$$

and the probability becomes

$$P(s_1) = Ae^{-(E(s_1)+\Delta\mu N_A(s_1))/k_BT}. \tag{4.25}$$

Furthermore, the probabilities of all states must sum up to 1

$$\sum_{s_i} P(s_i) = A\sum_{s_i} e^{-(E(s_i)+\Delta\mu N_A(s_1))/k_BT} = 1, \tag{4.26}$$

which leads to

$$A = \frac{1}{\sum_{s_i} e^{-(E(s_i)+\Delta\mu N_A(s_i)))/k_BT}}. \tag{4.27}$$

Thus we see that the normalization constant A is equal to $1/\mathcal{Z}$ where $\mathcal{Z}$ is the partition function and the probability of state s_1 becomes

$$P(s_1) = \frac{e^{-(E(s_1)+\Delta\mu N_A(s_1))/k_BT}}{\mathcal{Z}} = \frac{e^{-(E(s_1)+\Delta\mu N_A(s_1))/k_BT}}{\sum_{s_i} e^{-(E(s_i)+\Delta\mu N_A(s_i)))/k_BT}}. \tag{4.28}$$

4.6 The canonical ensemble

In the canonical ensemble the volume V, the temperature, T, and the number of atoms for each species N_i are fixed. Equation (4.28) is then simplified so that the probability of a state only depends on its energy,

$$P_{\text{canonical}}(s_1) = \frac{e^{-E(s_1)/k_BT}}{\sum_{s_i} e^{-E(s_i)/k_BT}}. \tag{4.29}$$

The probability ratio used in the MC simulation becomes

$$\frac{P_{\text{canonical}}(s_1)}{P_{\text{canonical}}(s_2)} = e^{(E(s_2)-E(s_1))/k_B T}. \tag{4.30}$$

Since the number of atoms of each kind is kept fixed the only configurations being explored in the MC simulations correspond to a re-ordering of atoms from the initial configuration. Therefore in lattice based models trial steps consist of swapping the species between two sites.

4.7 Semi-grand canonical ensemble

In the semi-grand canonical (SGC) ensemble the volume V, the temperature T, the chemical potential difference(s) $\Delta\mu_i$ and the number of sites N are fixed but the composition is allowed to change. Equation (4.28) is then kept as is and the probability of a state becomes (for a binary system)

$$P_{\text{SGC}}(s_1) = \frac{e^{-(E(s_1)+\Delta\mu N_A(s_1))/k_B T}}{\sum_{s_i} e^{-(E(s_i)+\Delta\mu N_A(s_i)))/k_B T}}. \tag{4.31}$$

The probability ratio used in the MC simulation becomes

$$\frac{P_{\text{SGC}}(s_1)}{P_{\text{SGC}}(s_2)} = e^{(E(s_2)-E(s_1)-\Delta\mu\Delta N_A)/k_B T}, \tag{4.32}$$

In the SGC ensemble the concentrations (yet not the total number of sites) are allowed to change. Therefore the trial step consists of selecting a site and changing its occupation to another species.

4.8 Validity of lattice based models

The partition function that has been discussed in this chapter only considers the summation over occupations $\boldsymbol{\sigma}$. The CE technique namely only considers the occupation vector and not displacements of the atoms. For real materials, however, the atoms also undergo vibrations around their equilibrium positions. This displacive degree of freedom is thus neglected in the CE, yet it can play a role in the actual value of a thermodynamical average. This might raise concerns about the validity of the calculated thermodynamical averages using CEs to represent the Hamiltonian in MC simulations. This section therefore considers the approximations made when using a CE in MC simulations for calculating thermodynamical averages.

In the canonical ensemble (Sect. 4.6) the partition function is defined as

$$\mathcal{Z}(N,V,T) = \sum_{\text{all states j}} \exp\left[-\beta E_j(N,V,T)\right]. \tag{4.33}$$

The partition function involves a sum over all microstates j of the system, each of which can be split up into one part consisting of the ordering of the atoms on a lattice and one part consisting of all possible displacements of the atoms for that specific ordering

$$\mathcal{Z}(N,V,T) = \sum_{\{\boldsymbol{\sigma}\}} \sum_{\{\nu \in \boldsymbol{\sigma}\}} \exp\left[-\beta E\left(\boldsymbol{\sigma},\nu,N,V,T\right)\right], \tag{4.34}$$

where $\{\boldsymbol{\sigma}\}$ is the set of all possible $\boldsymbol{\sigma}$ with constant N and $\{\nu \in \boldsymbol{\sigma}\}$ is the set of all displacements of the atoms which project on $\boldsymbol{\sigma}$. Thus $\{\nu \in \boldsymbol{\sigma}\}$ represents a subspace of the phase space of the original ensemble. The partition function can then be written as

$$\mathcal{Z}(N,V,T) = \sum_{\{\boldsymbol{\sigma}\}} \Lambda(\boldsymbol{\sigma},N,V,T), \tag{4.35}$$

with

$$\Lambda(\boldsymbol{\sigma},N,V,T) = \sum_{\{\nu \in \boldsymbol{\sigma}\}} \exp\left[-\beta E\left(\boldsymbol{\sigma},\nu,N,V,T\right)\right]. \tag{4.36}$$

In other words $\Lambda(\boldsymbol{\sigma},N,V,T)$ is the partition function for the subspace of the full ensemble for which all microstates project to the same configuration $\boldsymbol{\sigma}$. We can associate a free energy, $H(\boldsymbol{\sigma},N,V,T)$, to this partition function as

$$H(\boldsymbol{\sigma},N,V,T) = -\frac{1}{\beta}\ln\Lambda(\boldsymbol{\sigma},N,V,T). \tag{4.37}$$

The canonical partition function can then be written as

$$\mathcal{Z}(N,V,T) = \sum_{\{\boldsymbol{\sigma}\}} \exp\left(-\beta H(\boldsymbol{\sigma},N,V,T)\right). \tag{4.38}$$

Consider now the Hamiltonian $H(\boldsymbol{\sigma},\nu,N,V,T)$, which represents the free energy of a system, in which there are only displacive degrees of freedom. For a given $\boldsymbol{\sigma}$ it can be expressed as

$$H(\boldsymbol{\sigma},\nu,N,V,T) = U^0(\boldsymbol{\sigma},N,V,T) + F_{\text{ex}}(\boldsymbol{\sigma},\nu,N,V,T), \tag{4.39}$$

where U^0 is the energy of the static configuration in the $\{\nu \in \boldsymbol{\sigma}\}$ subspace with the lowest energy, i.e. the fully relaxed configuration and F_{ex} will then contain the remaining part of the free energy. The probability of a state described by $\boldsymbol{\sigma}$ is then

$$P(\boldsymbol{\sigma},N,V,T) = \frac{\exp\left(-\beta\left[U^0(\boldsymbol{\sigma},N,V,T) + F_{\text{ex}}(\boldsymbol{\sigma},N,V,T)\right]\right)}{\mathcal{Z}(N,V,T)}. \tag{4.40}$$

As it has been shown in this chapter, in order to use P for calculating thermodynamical averages one needs to evaluate ratios of probabilities P,

$$\frac{P(\boldsymbol{\sigma}_1, N, V, T)}{P(\boldsymbol{\sigma}_2, N, V, T)} = \frac{\exp\left(-\beta\left[U^0(\boldsymbol{\sigma}_1, N, V, T) + F_{\text{ex}}(\boldsymbol{\sigma}_1, N, V, T)\right]\right)}{\exp\left(-\beta\left[U^0(\boldsymbol{\sigma}_2, N, V, T) + F_{\text{ex}}(\boldsymbol{\sigma}_2, N, V, T)\right]\right)}. \tag{4.41}$$

$U^0\left(\boldsymbol{\sigma}, N, V, T\right)$ can now be approximated by taking the zero Kelvin value, which effectively is to approximate the Fermi-Dirac distribution by a step function [36]. Furthermore, in practice it is expected that F_{ex} is dominated by the vibrational free energy. A common approximation is therefore to remove the dependency of $\boldsymbol{\sigma}$ on F_{ex} and also assume that F_{ex} is a linear combination of the composition [36]

$$F_{\text{ex}}(N, V, T) = N_a F_{\text{ex},A}(T) + N_b F_{\text{ex},B}(T). \tag{4.42}$$

If F_{ex} is linear in the composition then it can be completely removed in the canonical ensemble since it cancels out in the probability ratio. For the SGC ensemble it can also for many purposes be left out completely since a linear combination of concentration is just a shift of the chemical potential by a constant. The value of $U^0(\boldsymbol{\sigma}, N, V)$ can now be calculated with a CE that maps $\boldsymbol{\sigma}$ to the fully relaxed configuration. Finally P reduces to

$$P(\boldsymbol{\sigma}, N, V, T) = \frac{1}{\mathcal{Z}} e^{-\beta U^0(\boldsymbol{\sigma}, N, V)}. \tag{4.43}$$

A thermodynamic average of a property $A(\boldsymbol{\sigma}, \nu)$ at N, V, T is then approximated by the value of A for the fully relaxed configuration. This means that even if the cluster expansion acts on the perfect lattice $\boldsymbol{\sigma}$, relaxation effects are still taken into consideration when calculating $\langle A \rangle$. Yet, any vibrational or temperature dependence of a property is largely neglected in this approach. If vibrational effects are important there are ways to incorporate these effects into a MC by also including the vibrational part of the free energy in the CE [37].

5

Summary of the papers

5.1 Paper I

Paper I focused on the inorganic clathrate $Ba_8Ga_{16}Ge_{30}$. A CE was constructed to represent the energies of fully relaxed structures obtained from DFT calculations and subsequently sampled by MC simulations, from which the chemical order as a function of temperature was obtained, specifically the SOFs. Representative configurations for specific temperatures were extracted from the simulations and further analyzed with respect to their electrical transport properties using DFT

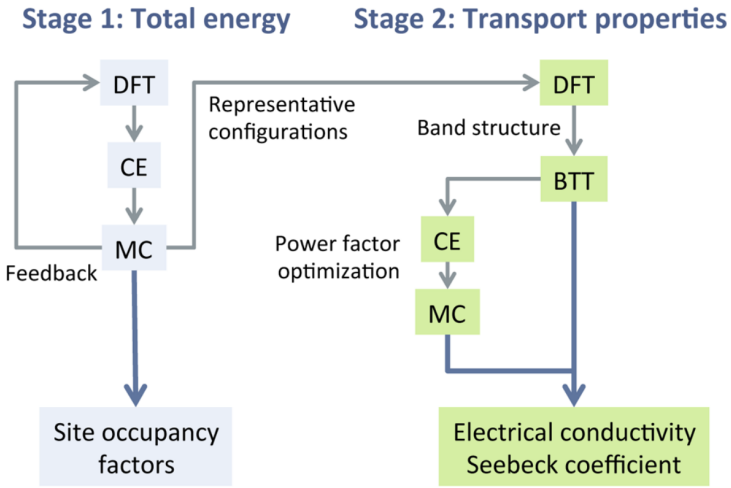

Figure 5.1: Schematic illustration of the methods employed in paper I. DFT: density functional theory; MC: Monte Carlo simulations; CE: alloy cluster expansion; BTT: Boltzmann transport theory.

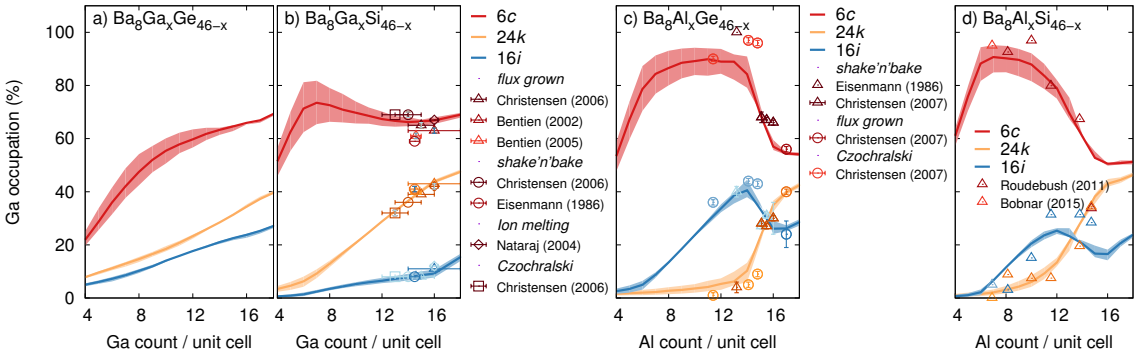

Figure 5.2: Site occupancy factors in intermetallic clathrates as a function of composition for Wyckoff sites 6*c* (red), 16*i* (blue), and 24*k* (orange). Solid lines show simulation results obtained at 700 K whereas the shaded regions indicate a variation by ± 100 K.

and Boltzmann transport theory (BTT) calculations. This combination of CE and MC simulations as well as DFT and BTT allowed us to obtain the thermoelectric power factor as a function of temperature (Fig. 5.1). The approach was validated by comparing the SOFs and transport coefficients with experimental data. Another CE was then constructed to represent the power factor as a function of the chemical configuration and was employed to determine the chemical ordering that maximized the power factor. The optimized structure yielded a power factor increase by more than 60% that was achieved by reducing the number of trivalent species on the 6*c* Wyckoff site. Hence, the approach developed in this paper demonstrates the use of CEs for structure optimization.

5.2 Paper II

Paper II addressed the chemical ordering in the clathrate systems $Ba_8Ga_xGe_{46-x}$, $Ba_8Ga_xSi_{46-x}$, $Ba_8Al_xGe_{46-x}$, and $Ba_8Al_xSi_{46-x}$ as a function of composition (Fig. 5.2). In particular Al-based clathrates were found to display an extreme variation of SOFs as function of composition (Fig. 5.2c,d). The ordering in these materials can significantly impact the material properties, see e.g., paper I. Hence, an understanding of the ordering is crucial in order to understand and optimize these materials. To that end, in paper II for each system a CE was constructed based on the energies of fully relaxed structures obtained from DFT calculations and the SOFs were obtained from MC simulations. The simulated SOFs agree very well with experimental data (Fig. 5.2), which allowed us to clarify variations and trends in the experimental data. In particular the CE-MC simulations provide an explanation of the extreme variations of the SOFs in Al-based clathrates.

Acknowledgments

To my family, thank you for all your love and support.

To Emma, thank you for making our home such a warm and affectionate place.

To my friends, both local and those from up north, thank you for all the fun times.

To my supervisor Paul Erhart, thank you for giving me the opportunity to do this work. Your dedication and knowledge about physics and especially your enthusiasm to share it is greatly appreciated. Also, thanks for your unwavering demand of quality in all of my work.

To my examiner, Göran Wahnström, thank you for your wisdom and leadership. Also, thank you for your dedication to deadlines which made sure this thesis was written on time.

I would also like to thank Thomas Holm Rod and the whole data analysis & modelling group at ESS in Copenhagen for welcoming me to visit the group and learn how professional software development is carried out.

To my office mate, Erik Fransson, thank you for being such a good friend and rubber duck. I only want to say one thing: *Tänk dig att du har en cell....*

To William Armando Muñoz, Magnus Rahm, Erik Fransson and Paul Erhart, thank you for working with me to develop icet.

To Joakim Brorsson and Anders Palmqvist, thank you for many helpful discussions regarding clathrates.

Finally, I am grateful for all current and former members of the Materials and Surface theory division that provide such a fun and exciting work atmosphere.

Bibliography

[1] T. Caillat, J.-P. Fleurial, and A. Borshchevsky, *Preparation and thermoelectric properties of semiconducting Zn_4Sb_3*, Journal of Physics and Chemistry of Solids **58**, 1119 (1997).

[2] E. S. Toberer, K. A. Sasaki, C. R. I. Chisholm, S. M. Haile, G. W. A., and S. G. J., *Local structure of interstitial Zn in β-Zn_4Sb_3*, physica status solidi (RRL) – Rapid Research Letters **1**, 253 (2007).

[3] G. Jeffrey Snyder, M. Christensen, E. Nishibori, T. Caillat, and B. Brummerstedt Iversen, *Disordered zinc in Zn_4Sb_3 with phonon-glass and electron-crystal thermoelectric properties*, **3**, 458 (2004).

[4] H. J. Kim, E. S. Božin, S. M. Haile, G. J. Snyder, and S. J. L. Billinge, *Nanoscale α-structural domains in the phonon-glass thermoelectric material β—Zn_4Sb_3*, Phys. Rev. B **75**, 134103 (2007).

[5] C. Uher, in *Recent Trends in Thermoelectric Materials Research I*, Vol. 69 of *Semiconductors and Semimetals*, edited by T. M. Tritt (Elsevier, Atlanta, 2001), pp. 139 – 253.

[6] B. C. Sales, D. Mandrus, and R. K. Williams, *Filled Skutterudite Antimonides: A New Class of Thermoelectric Materials*, Science **272**, 1325 (1996).

[7] G. P. Meisner, D. T. Morelli, S. Hu, J. Yang, and C. Uher, *Structure and Lattice Thermal Conductivity of Fractionally Filled Skutterudites: Solid Solutions of Fully Filled and Unfilled End Members*, Phys. Rev. Lett. **80**, 3551 (1998).

[8] G. S. Nolas, J. L. Cohn, and G. A. Slack, *Effect of partial void filling on the lattice thermal conductivity of skutterudites*, Phys. Rev. B **58**, 164 (1998).

[9] G. Snyder and E. Toberer, *Complex thermoelectric materials*, Nature Materials **7**, 105 (2008).

[10] D. T. Morelli and G. P. Meisner, *Low temperature properties of the filled skutterudite* $CeFe_4Sb_{12}$, Journal of Applied Physics **77**, 3777 (1995).

[11] B. C. Sales, D. Mandrus, B. C. Chakoumakos, V. Keppens, and J. R. Thompson, *Filled skutterudite antimonides: Electron crystals and phonon glasses*, Phys. Rev. B **56**, 15081 (1997).

[12] H. Kim, M. Kaviany, J. C. Thomas, A. Van der Ven, C. Uher, and B. Huang, *Structural Order-Disorder Transitions and Phonon Conductivity of Partially Filled Skutterudites*, Phys. Rev. Lett. **105**, 265901 (2010).

[13] M. Christensen, S. Johnsen, and B. B. Iversen, *Thermoelectric clathrates of type I*, Dalton Trans. **39**, 978 (2010).

[14] J. H. Roudebush, C. de la Cruz, B. C. Chakoumakos, and S. M. Kauzlarich, *Neutron Diffraction Study of the Type I Clathrate $Ba_8Al_xSi_{46-x}$: Site Occupancies, Cage Volumes, and the Interaction between the Guest and the Host Framework*, Inorg. Chem. **51**, 1805 (2012).

[15] M. Ångqvist and P. Erhart, *Understanding Chemical Ordering in Intermetallic Clathrates from Atomic Scale Simulations*, Chemistry of Materials **29**, 7554 (2017).

[16] M. Ångqvist, D. O. Lindroth, and P. Erhart, *Optimization of the Thermoelectric Power Factor: Coupling between Chemical Order and Transport Properties*, Chem. Mater. **28**, 6877 (2016).

[17] J. Walecka, *Introduction to Statistical Mechanics* (World Scientific, 5 Toh Tuck Link, Singapore 596224, 2011).

[18] D. Porter, K. Easterling, and M. Sherif, *Phase Transformations in Metals and Alloys, Third Edition (Revised Reprint)* (CRC Press, 6000 broken sound parkway nw suite 300, 2009).

[19] L. Gharaee, M. Ångqvist, M. Rahm, and P. Erhart (unpublished).

[20] P. Rogl, *Thermoelectrics Handbook* (CRC Press, 6000 broken sound parkway nw suite 300, 2005), Chap. 32, pp. 1–24.

[21] A. V. Shevelkov and K. Kovnir, in *Zintl Phases*, No. 139 in *Structure and Bonding*, edited by T. F. Fässler (Springer Berlin Heidelberg, Tiergartenstraße 17, 69121 Heidelberg, 2011), pp. 97–142.

[22] M. A. Avila, K. Suekuni, K. Umeo, H. Fukuoka, S. Yamanaka, and T. Takabatake, *$Ba_8Ga_{16}Sn_{30}$ with type-I clathrate structure: Drastic suppression of heat conduction*, Applied Physics Letters **92**, 041901 (2008).

[23] G. S. Nolas, J. L. Cohn, G. A. Slack, and S. B. Schujman, *Semiconducting Ge clathrates: Promising candidates for thermoelectric applications*, Appl. Phys. Lett. **73**, 178 (1998).

[24] G. K. H. Madsen, K. Schwarz, P. Blaha, and D. J. Singh, *Electronic structure and transport in type-I and type-VIII clathrates containing strontium, barium, and europium*, Phys. Rev. B **68**, 125212 (2003).

[25] T. Matsui, J. Furukawa, K. Tsukamoto, H. Tsuda, and K. Morii, *Structure and properties of $Ba_8Ga_{16}Ge_{30}$ clathrates by a novel synthesis method using CO gas reductive atmosphere*, Journal of Alloys and Compounds **391**, 284 (2005).

[26] N. P. Blake, S. Latturner, J. D. Bryan, G. D. Stucky, and H. Metiu, *Band structures and thermoelectric properties of the clathrates $Ba_8Ga_{16}Ge_{30}$, $Sr_8Ga_{16}Ge_{30}$, $Ba_8Ga_{16}Si_{30}$, and $Ba_8In_{16}Sn_{30}$*, The Journal of Chemical Physics **115**, 8060 (2001).

[27] M. Troppenz, S. Rigamonti, and C. Draxl, *Predicting Ground-State Configurations and Electronic Properties of the Thermoelectric Clathrates $Ba_8Al_xSi_{46-x}$ and $Sr_8Al_xSi_{46-x}$*, Chem. Mater. **29**, 2414 (2017).

[28] J. M. Sanchez, F. Ducastelle, and D. Gratias, *Generalized cluster description of multicomponent systems*, Physica **128**, 334 (1984).

[29] A. van de Walle, *Multicomponent multisublattice alloys, nonconfigurational entropy and other additions to the Alloy Theoretic Automated Toolkit*, Calphad **33**, 266 (2009), tools for Computational Thermodynamics.

[30] L. J. Nelson, G. L. W. Hart, F. Zhou, and V. Ozoliņš, *Compressive sensing as a paradigm for building physics models*, Phys. Rev. B **87**, 035125 (2013).

[31] L. J. Nelson, V. Ozoliņš, C. S. Reese, F. Zhou, and G. L. W. Hart, *Cluster expansion made easy with Bayesian compressive sensing*, Phys. Rev. B **88**, 155105 (2013).

[32] E. Candes and M. Wakin, *An Introduction To Compressive Sampling*, IEEE Signal Processing Magazine **25**, 21 (2008).

[33] T. Goldstein and S. Osher, *The Split Bregman Method for L1-Regularized Problems*, SIAM Journal on Imaging Sciences **2**, 21 (2009).

[34] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, *Bregman Iterative Algorithms for l_1-Minimization with Applications to Compressed Sensing*, SIAM Journal on Imaging Sciences **1**, 26 (2008).

[35] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equation of State Calculations by Fast Computing Machines*, The Journal of Chemical Physics **21**, 1087 (1953).

[36] G. Ceder, *A derivation of the Ising model for the computation of phase diagrams*, Computational Materials Science **1**, 144 (1993).

[37] A. van de Walle and G. Ceder, *The effect of lattice vibrations on substitutional alloy thermodynamics*, Rev. Mod. Phys. **74**, 11 (2002).